Exhibit M

IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF VIRGINIA
ALEXANDRIA DIVISION

United States of America, *et al.*,

Plaintiffs,

v

Google LLC,

Defendant.

Case No. 1:23-cv-00108

HON. LEONIE H. M. BRINKEMA

EXPERT REPORT OF

GABRIEL WEINTRAUB, PH.D.

DECEMBER 22, 2023

[redacted][247] Given the need for "sufficient" and the "right kind" of data for these algorithms, as outlined in this subsection, a company that falls out of the virtuous cycle will likely face difficulty in overcoming the hurdle when data scale, auction outcomes such as impressions won and revenue, and thickness are intertwined.

III.C. Scale and Experimentation

100. Technology companies routinely use experiments to assess the impact of changing product (or service) features or introducing new features. For these experiments, they often use A/B tests, also known as online controlled experiments ("OCE").[248] An A/B test is a randomized experiment in which a fraction of subjects are randomly selected to receive a treatment while the remaining subjects form the control group that represents the baseline (which is often the status quo). [redacted][249] [redacted][250] [redacted]

---

247 [redacted] *see also*, Deposition of John Gentry (OpenX), October 26, 2023, 63:17–64:13 ("Q. Does OpenX vary its take rate by impression, meaning for a given publisher, OpenX might charge a different take rate for one impression than the other?…THE WITNESS: Yes. It varies – that varies by publisher in terms of how we apply our take rate, but a simple example is that we will often do something called fee squashing where we'll take a lower fee if we need to, to try and get an impression to run through our system. That would be an example.…Q. How, if at all, does greater scale affect exchanges' ability to effectively set different take rates for different impressions? A.…The greater the scale that you have, the more data you have. The more data you have, you can better decide what an appropriate fee would be on a given impression.").

248 Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, (Cambridge University Press, 2020), 5 ("Controlled experiments have a long and fascinating history, which we share online (Kohavi, Tang and Xu 2019). They are sometimes called A/B tests, A/B/n tests (to emphasize multiple variants), field experiments, randomized controlled experiments, split tests, bucket tests, and flights. In this book, we use the terms *controlled experiments* and *A/B tests* interchangeably, regardless of the number of variants.") (emphasis in original).

249 [redacted]

250 [redacted]

[redacted][251]
[redacted][252]
[redacted]
[redacted]
[redacted][253]

101. A/B tests are considered to be a scientifically valid approach to determine the causal impact of a treatment.[254] These randomized OCE's are the "gold standard for establishing causality,"[255] which is why web-facing companies including eBay, Facebook, Google, and Microsoft (among

---

[redacted]

251 [redacted]

252 [redacted]

253 [redacted]

254 Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge: Cambridge University Press, 2020), 9 ("Randomized controlled experiments are the gold standard for establishing causality.").

255 Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 9 ("Randomized controlled experiments are the gold standard for establishing causality.").

many others), use them to guide product development and accelerate innovation.[256] When designing experiments, there are best practices to ensure that the results are statistically valid;

[redacted][257]

[redacted]

[redacted][258]

[redacted]

[redacted][259]

---

256 Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann, "Online Controlled Experiments at Large Scale," *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013): 1–9, at 1 ("Many web-facing companies use online controlled experiments to guide product development and prioritize ideas, including Amazon [1], eBay, Etsy [2], Facebook, Google [3], Groupon, Intuit [4], LinkedIn, Microsoft [5], Netflix [6], Shop Direct [7], StumbleUpon [8], Yahoo, and Zynga [9].").

257 [redacted]

258 [redacted]

[redacted][260][redacted]

102. In this way, A/B tests have become an integral part of technological companies' innovation engines. These companies rely on experiments to gauge the effects of potential changes to their platforms before implementation.[261] Organizations use these experiments as a key tool for developing frameworks and testing ideas to understand the impact of the services they provide;[262] they have become a standard component of managerial decision making.[263] Companies like eBay, Facebook, Microsoft and Google run thousands to tens of thousands of experiments every year, testing everything from changes to the user interface, to customer

---

[redacted]

260

261 Michael Luca, and Max H. Bazerman, *The Power of Experiments: Decision Making in a Data-Driven World* (Cambridge: MIT Press, 2021): vii-viii ("These days, companies like Google wouldn't dare make a major change in their platforms without first looking at experiments to understand how it would influence user behavior.").

262 Michael Luca and Max H. Bazerman, *The Power of Experiments: Decision Making in a Data-Driven World* (Cambridge: MIT Press, 2021), vii–viii ("From startups to international conglomerates to government agencies, organizations have a new tool to develop frameworks and test ideas, and to understand the impact of the products and services they are providing.").

263 Michael Luca and Max H. Bazerman, *The Power of Experiments: Decision Making in a Data-Driven World* (Cambridge: MIT Press, 2021), 62 ("But perhaps no sector has embraced the experimental method more than the tech sector, where it is now a standard component of managerial decision making.").

support systems, across websites, mobile apps and more.[264] [redacted][265]

103. A/B tests are often conducted with a relatively small yet representative sample [redacted][266] In this way, platforms reduce the potential risk of having a new feature produce an unexpected impact. For example, the new feature may degrade users' experiences, may reduce users' performance, or may involve an unexpected economic cost for the platform. The existence or magnitude of these negative effects may be unknown in advance; hence, a way of mitigating risk is to treat a relatively small fractions of units.

104. Statistical power is an important criterion to consider when conducting statistically sound experiments.[267] Suppose an experiment is conducted to detect an X percent (or larger) improvement in an outcome of interest caused by a new feature. This "X percent" could be the smallest detectable effect that the company would use as a threshold to decide whether to launch the feature system-wide. Statistical power refers to the probability of detecting such an effect

---

[264] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 5 ("Online controlled experiments are used heavily at companies like Airbnb, Amazon, Booking.com, eBay, Facebook, Google, LinkedIn, Lyft, Microsoft, Netflix, Twitter, Uber, Yahoo!/Oath, and Yandex (Gupta et al. 2019). These companies run thousands to tens of thousands of experiments every year, sometimes involving millions of users and testing everything, including changes to the user interface (UI), relevance algorithms (search, ads, personalization, recommendations, and so on), latency/performance, content management systems, customer support systems, and more. Experiments are run on multiple channels: websites, desktop applications, mobile applications, and e-mail.").

[265] [redacted]

[266] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, (Cambridge University Press, 2020): 171 ("In practice, it is common that an experiment goes through a *ramping* process to control unknown risks associated with new feature launches (aka. controlled exposure). For example, a new feature may start by exposing the Treatment to only a small percentage of users. If the metrics look reasonable and the system scales well, then we can expose more and more users to the Treatment. We ramp the traffic until the Treatment reaches desired exposure level.") [redacted]

[267] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("Practically speaking, you want enough power in your experiment to be able to conclude with high probability whether your experiment has resulted in a change bigger than what you care about.").

when in fact one exists.[268] There are multiple factors that affect the statistical power of an experiment, such as the number of observations, where fewer observations lead to lower statistical power, and variance in the outcome of interest, where higher variance leads to lower statistical power.[269]

105. A significant challenge in running A/B tests for technology platforms that run many experiments simultaneously is to secure enough users to achieve a desired level of statistical power. Technology companies are constantly launching, changing, and iterating on their products.[270] Each individual change may have a small effect, but the compounding effect of many small changes can be significant.[271] Therefore, detecting small effects reliably (or in a statistically significant way) is critical to be able to continuously experiment and improve.

106. Statistical theory shows the size of the experimental sample should become larger as the level of the effect that an experimenter wants to detect becomes smaller.[272] For example, if a company

---

[268] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("*Statistical power* is the probability of detecting a meaningful difference between the variants when there really is one (statistically, reject the null when there is a difference).") (emphasis in original).

[269] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("Usually, we get more power when the sample size is larger."), and 189–190 ("Assuming Treatment and Control are of equal size, the total number of samples you need to achieve 80% power can be derived from the power formula above, and is approximately as showing in Equation 17.8 (van Belle 2008): $n \approx \frac{16\sigma^2}{\delta^2}$ where, σ^2 is the sample variance, and δ is the difference between Treatment and Control.").

[270] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 13 ("Features are built because teams believe they are useful, yet in many domains most ideas fail to improve key metrics. Only one third of the ideas tested at Microsoft improved the metric(s) they were designed to improve (Kohavi, Crook and Longbotham 2009). Success is even harder to find in well-optimized domains like Bing and Google, whereby some measures' success rate is about 10-20% (Manzi 2012).").

[271] Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 14 ("[Google] ran hundreds of controlled experiments and multiple iterations; some across all markets, and some long term in specific markets to understand the impact on advertisers in more depth. This large backend change – and running controlled experiments – ultimately validated how planning multiple changes and layering them together improved the user's experience by providing higher quality ads, and improved their advertiser's experience moving towards lower average prices for the higher quality ads.").

[272] Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann, "Online Controlled Experiments at Large Scale," *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2013): 1–9, at 1 ("While running online controlled experiments require a sufficient number of users, teams working on products with thousands to tens of thousands of users (our general guidance is at least thousands of active users) are typically looking for larger effects, which are easier to detect than the

does not want to introduce a new feature unless an expected effect has a 1 percent lift on traffic, it should run a larger sample size experiment than if their target effect is 2 percent. As a result, large sample sizes are required for conducting reliable A/B tests for individual changes with relatively small effects.[273]

107. For this reason, it can be challenging for firms with limited scale to run informative experiments.[274] For technology platforms, the unit of an observation in an experiment typically corresponds to a user using the platform (e.g., a bidder), or to an event (e.g., a query, an auction or an impression). Hence, typically platforms with smaller scale need to run experiments over a longer time period before the number of users or events included in the experiment reaches the sample size required to detect the desired effect.[275] Running such long experiments can be impractical when a firm's objective is to iterate many small changes quickly. For example, most large tech companies run a single experiment for two weeks or less.[276] Consider another smaller company that has 20 percent of the traffic of the large tech company and uses the same treatment allocation fraction as the large company. Then, everything else equal, this smaller company

---

small effects that large sites worry about. For example, to increase the experiment sensitivity (detectable effect size) by a factor of 10, say from 5% to 0.5%, you need $10^2 = 100$ times more users.").

273 Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 30 ("Usually, we get more power when the sample size is larger.").

274 [redacted] *see also*, "Find top products in Learning Management Systems (LMS) category," LinkedIn, accessed December 20, 2023, https://www.linkedin.com/products/categories/learning-management-systems.

275 Ron Kohavi, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing* (Cambridge University Press, 2020), 33 ("**More users:** In the online experiments, because users trickle into experiments over time, the longer the experiment runs, the more users the experiment gets. This usually results in increased statistical power") (emphasis in original); [redacted]

276 Somit Gupta, Ronny Kohavi, Diane Tang, Ya Xu, et al., "Top Challenges from the First Practical Online Controlled Experiments Summit," *ACM SIGKDD Explorations Newsletter* 21, no. 1 (2019): 20–35, at 21–22 ("There are many interesting and challenging open questions for OCE results analysis. While most experiments in the industry run for 2 weeks or less, we are really interested in detecting the long-term effect of a change… At Microsoft, while most experiments do not run for more than two weeks, it is recommended to run an experiment longer if novelty effects are suspected and use data from the last week to estimate the long-term treatment effect…At Twitter, a similar practice is followed. An experiment at Twitter may run for 4 weeks and data fromlast two weeks is analyzed.").

Gabriel Weintraub, Ph.D.

December 22, 2023

[redacted]

Expert Report of Gabriel Weintraub (December 22, 2023)--Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
7	16	6	"ad.spend"	"ad spend"	Typo
[redacted]					
11	22	10	"See, e.g., "Internet Advertising Revenue Report," PwC, iab, April 2023, https://www.iab.com/wpcontent/uploads/2023/04/IAB_PwC_Internet_Advertising_Revenue_Report_2022.pdf"	"See, e.g., "Internet Advertising Revenue Report," PwC, iab, April 2023, https://www.iab.com/wpcontent/uploads/2023/04/IAB_PwC_Internet_Advertising_Revenue_Report_2022.pdf, **at 15**"	Clarification
12	25	.	"The tools to run RTB auctions for the purchase and sale of online display ads **is** called…"	"The tools to run RTB auctions for the purchase and sale of online display ads **are** called…"	Typo
17	33	41	"(e.g., the average spend for each conversion should not exceed a pre-specified target).")."	"(e.g., the average spend for each conversion should not exceed a pre-specified target).") **(emphasis in original)**."	Clarification
19	34	47	"at 39**68**"	"at 39**52**"	Correction
23	43	71	"using machine algorithms"	"using machine **learning** algorithms"	Clarification
24	44	77	"no. 1 (**2014**)"	"no. 1 (**2015**)"	Typo
25	46	88	"EC '22: Proceedings of the **2023** ACM Conference on Economics and Computation **2023**"	"EC '22: Proceedings of the **23rd** ACM Conference on Economics and Computation"	Typo
25	46	88	"no. 1 (**2023**)"	"no. 1 (**2022**)"	Typo
25	46	90	"EC '22: Proceedings of the **2023** ACM Conference on Economics and Computation **2023**"	"EC '22: Proceedings of the **23rd** ACM Conference on Economics and Computation"	Typo
25	46	90	"no. 1 (**2023**)"	"no. 1 (**2022**)"	Typo
26	46	90	"(2023): 1–35, at 6"	"(2023): 1–35, at 6**–7**"	Clarification
26	46	91	"EC '22: Proceedings of the **2023** ACM Conference on Economics and Computation **2023**"	"EC '22: Proceedings of the **23rd** ACM Conference on Economics and Computation"	Typo
26	46	91	"no. 1 (**2023**)"	"no. 1 (**2022**)"	Typo
[redacted]					
28	51	.	"Google Ad Manager ("GAM"), which includes a publisher ad **server, and** an ad exchange/SSP."	"Google Ad Manager ("GAM"), which includes a publisher ad **server and** an ad exchange/SSP."	Typo
32	59	132	"$\theta^1(x)$ "	"$\theta^1(x)$ "	Typo
36	68	143	"("we use the *market thickness* to represent the average number of ads competing for user impressions on an online advertising platform.") **(emphasis in original)**;"	"("we use the *market thickness* to represent the average number of ads competing for user impressions on an online advertising platform.");"	Correction
[redacted]					
38	70	154	"**955**–1025"	"**965**–1025"	Typo
38	70	157	"**955**–1025"	"**965**–1025"	Typo
[redacted]					
56	93	225	"1849–1864, at **1850**"	"1849–1864, at **1851**"	Typo
60	98	243	"209–218, at **217**"	"209–218, at **209**"	Correction
60	98	244	"209–218, at **217**"	"209–218, at **216**"	Correction
[redacted]					
64	102	261	"(Cambridge: MIT Press, **2021**)"	"(Cambridge: MIT Press, **2020**)"	Typo
64	102	262	"(Cambridge: MIT Press, **2021**)"	"(Cambridge: MIT Press, **2020**)"	Typo
64	102	263	"(Cambridge: MIT Press, **2021**)"	"(Cambridge: MIT Press, **2020**)"	Typo
66	104	269	"and 189**–190**"	"and 189"	Clarification

Expert Report of Gabriel Weintraub (December 22, 2023)--Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
68	109	279	"at 3933"	"at 3933**–3934**"	Clarification
68	109	279	"(2017): 500–522."	"(2017): 500–522**, at 500, 510**."	Clarification
70	112	284	"**23:6**–24:6"	"**24:1**–24:6"	Clarification
70	112	.	"John Gentry of OpenX named personnel costs alongside infrastructure costs, and real estate costs as costs that need **to covered** in order to have the "ability to invest and grow the business.""	"John Gentry of OpenX named personnel costs alongside infrastructure costs and real estate costs as costs that need **to be covered** in order to have the "ability to invest and grow the business.""	Typo
[redacted]					
73	115	296	"247:17–248:**19**"	"247:17–248:**7**"	Clarification
[redacted]					
115	169	396	"189:5–190:**18**"	"189:5–190:**3**"	Clarification
[redacted]					
146	220	.	"Recall that in first price auctions, it is optimal for bidders to reduce their bids (Section **II.A.1**)."	"Recall that in first price auctions, it is optimal for bidders to reduce their bids (Section **II.C.1**)."	Typo
[redacted]					
152	229	552	"Deposition of John Gentry (OpenX), October 26, 2023, 21:15–21:22 ("Q. How, if at all, did the decrease in spending by DV360 on OpenX affect OpenX?…THE WITNESS: It was a devastating impact to the company, resulting in severe financial consequences. We had to execute a large layoff in December of 2018 and had a lot of negative effects as a result of that.")."	Switch with text in footnote 553.	Correction

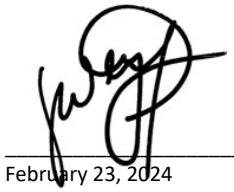
Expert Report of Gabriel Weintraub (December 22, 2023)--Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
152	229	553	"Deposition of John Gentry (OpenX), October 26, 2023, 22:17–23 ("Our revenues went from - we were about [redacted] in 2017; 2018 we were about [redacted] because the first half of the year was strong; in 2019, we were down to about [redacted]; and in 2020, the combination of difficulties we had had plus COVID took us down to about [redacted] in net revenue in 2020.")."	Switch with text in footnote 552.	Correction
152	229	.	"The CEO of OpenX, John Gentry, testified that OpenX's revenues fell from [redacted] in 2017 to [redacted] 2019." [redacted]	"The CEO of OpenX, John Gentry, testified that OpenX's revenues fell from [redacted] in 2017 to [redacted] 2019." [redacted]	Typo
C-1	3.a.ii.	3	"Letter from David R. Pearl to **Kelly Garcia**, September 8, 2023, 2."	"Letter from David Pearl to **Michael Freeman**, September 8, 2023, 2."	Correction
D-10	22	17	"Letter from Julie Elmer to John Hogan, August 19, 2022, 5–6**|8** (" [redacted]	"Letter from Julie Elmer to John Hogan, August 19, 2022, 5–6 (" [redacted]	Typo
E-2	2	.	"exchanges' bids **have the same variance** that is characterized **by the spread parameter**"	"exchanges' bids **have variance** that is characterized **by the same spread parameter**"	Clarification
E-21	42	21	"Letter from David Pearl to Michael J. Freeman, July 28, 2023 ("	"Letter from David Pearl to Michael J. Freeman, July 28, 2023**, 2** ("	Clarification
F-2	3	5	"John Rice, "11.2 Comparing Two Independent Samples", in Mathematical Statistics and Data Analysis, 3rd ed. (Duxbury: Thomson Brooks/Cole, 2007)."	"John Rice, "11.2 Comparing Two Independent Samples", in Mathematical Statistics and Data Analysis, 3rd ed. (Duxbury: Thomson Brooks/Cole, 2007)**, 421–444**."	Clarification
F-5	11	12	("Rule of Thumb[:] The basic formula is n=16/Δ^2 (2.3).").	("Rule of Thumb[:] The basic formula is n=16/Δ^2 (2.3) **where Δ=(μ_0-μ_1)/σ=δ/σ (2.4)**.").	Clarification
F-5	11	12	N/2=16 (p_p (1-p_p))/(p_1-p_0)	N/2=16 (p_p (1-p_p))/(p_1-p_0)**^2**	Correction

January 13, 2024

Expert Report of Gabriel Weintraub (December 22, 2023)--Supplemental Errata

Page	Paragraph	Footnote	Original	Corrected	Reason
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]



February 23, 2024